Computer Assisted Sphere Packing in Higher Dimensions

Nelson Max

University of California, Davis and
Lawrence Livermore National Laboratory

Abstract

A computer was used to help study the packing of equal spheres in dimension four and higher. A candidate for the densest packing in 4-space is described. The configuration of 24 spheres touching a central sphere in this packing is shown to be rigid, unlike the analogue in 3-space, in which the spheres can slide past each other. A system for interactively manipulating and visualizing such configurations is described.

The Voronoi cell for a sphere is the set of points closer to its center than to any other sphere center in the packing. The packing density is the ratio of a sphere's volume to the average of the volumes of the Voronoi cells. A method of constructing Voronoi cells and computing their volumes is presented, which works in any dimension. Examples of Voronoi cell volumes are given.

Introduction

Hsiang[1] has recently announced a proof that the face-centered cubic (FCC) packing has the highest density of any packing of equal spheres. (Actually it is in a tie with infinitely many other related packings.) A central sphere is surrounded by 12 spheres, each touching the central one and four of the 12 others. This packing is important in physics and chemistry, since it is common in crystals, and the highest density would he favored at high pressure. Kepler conjectured in 1611 that the FCC packing is the densest, but a mathematical proof has only been given this year. It is an interesting mathematical problem to find the densest packing in higher dimensions as well. As discussed by Conway and Sloane [2], there is a relationship between packings in higher dimensions, arrangements of points on the unit sphere, and error detecting and correcting codes for transmitting information. An easy way to describe the FCC packing is to put a sphere center at all points (i, j, k) on the integer lattice such that $i + j + k$ is even, and set the radius to .5 Sqrt(2).

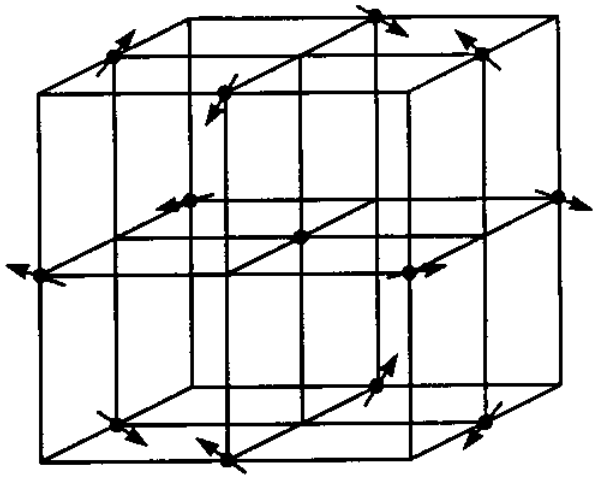

Figure 1. The FCC packing with velocity vectors.

In figure 1, the central dot is a sphere center at the origin (0, 0, 0). The other dots indicate sphere centers (i, j, k) at the midpoints of the 12 edges of a cube of side 2, where one of i, j, or k is zero, and the other two are +1 or -1. Thus a layer of the packing in the $i = 0$ plane is a tilted square array in which each sphere touches its four diagonal neighbors, and the layers at $i = 1$ or $i = -1$ fit into the gaps at the center of each square. There are also layers of hexagonally packed spheres in planes perpendicular to the four body diagonal vectors like (1, -1, 1). There is an analogous sphere packing, called D4, in 4- space, with sphere centers at all integer lattice points (i, j, k, l) such that $i + j + k + l$ is even. In this packing, a central sphere is surrounded by 24 others, each touching it and eight of the 24 others. Hsiang has conjectured that D4 is the densest packing in 4-space. To aid in studying this conjecture, he has asked four questions which a computer can help answer.

a) How many equal spheres can touch a central one? (A collection of equal non-overlapping spheres all tangent to a central one are called "kissing" spheres.)

b) Can the 24 kissing spheres in D4 slide past each other?

c) Given a collection of points in any number of dimensions, all surrounding a central point P, what is the volume of the "Voronoi cell" of P (containing all positions which are closer to P than to any other point in the collection)?

d) Does the D4 packing give the smallest possible Voronoi cell for kissing spheres?

Problem c) is useful in studying the packing density, since space can be divided up into the Voronoi cells corresponding to the sphere centers. The packing density is then the ratio of the volume of a sphere to the average of the volumes of the Voronoi cells.

The problem of maximum packing density in 3-space is complicated by the fact that the 12 kissing spheres in the FCC packing can slide past each other. Figure 1 shows arrows indicating the initial velocity vectors of the 12 kissing sphere centers during the sliding motion. On each square face of the cube, two opposite centers move toward each other, while the two others move farther away. This changes the convex hull of the sphere centers from a polyhedron bounded by 6 squares and 8 triangles into one bounded by 20 triangles. The spheres are now free to rattle. As explained by Conway and Sloane [2], they can move, for example, to kiss at the centers of the faces of a regular dodecahedron circumscribed about the central sphere.

The dodecahedron is the Voronoi cell for the central sphere in this configuration, and it has smaller volume than the Voronoi cell for the FCC packing. Hsiang [1] has shown that this configuration has the greatest possible "local density" (ratio of a single sphere's volume to that of its own Voronoi cell.) But this configuration does not extend to an efficient global packing, because the 12 surrounding spheres have gaps between them, making their Voronoi cells larger. In the proof announced in [1], it is necessary to study the possible configurations of up to 57 spheres in two layers surrounding a central one, in order to analyze the trade off between the Voronoi cell of the central sphere and those of the spheres in the first surrounding layer.

Hsiang hopes to avoid this complexity in 4-space by proving that the Voronoi cell for a sphere in the D4 packing has the smallest possible volume. Since every sphere in the D4 packing is equivalent, this would show that D4 gives the densest packing. Answers to questions a), b), and d) would be steps toward this proof.

Kissing Spheres (question a)

Spheres can easily be manipulated mentally or physically in 3-space, but this is much harder to do in 4-space. Therefore a model of the kissing spheres is manipulated in a computer, which continually checks whether they overlap. Visualization is still useful, so the spheres in 4-space are projected onto the 2-D screen using a combination of radial, stereographic, and parallel projections. The kissing spheres are first radially projected onto the surface of the central sphere, giving a collection of non-overlapping spheres of one lower dimension. This is illustrated in the 3-D case in figure 2. The central sphere is shown at the bottom, and above it are two kissing spheres, tangent to each other. Each projects radially onto the center sphere as a circular region, shown shaded in the figure. The union of the projection rays for each sphere form a cone with vertex at the center of the radial projection, and these cones intersect the central sphere in the shaded regions. Since the two kissing spheres are tangent, so are the cones, and therefore so are the shaded regions. In one dimension higher, the shaded regions would correspond to non-overlapping spheres on the 3-D surface of the central sphere in 4-space. The positions of these projected regions uniquely specify the positions of the kissing spheres, so visualizing them can help in analyze the kissing spheres. But it is still difficult to visualize this 3-D surface, so the next step is to project it onto ordinary 3-D Euclidean space.

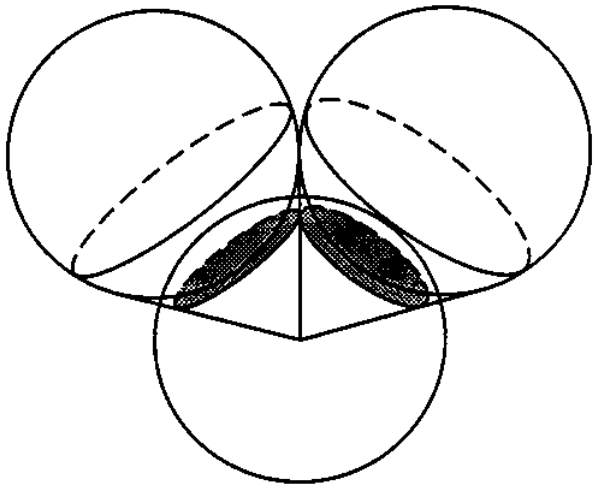

Figure 2. Two spheres projected onto a third sphere.

This is accomplished by stereographic projection, which works in any dimension. In 3-D, it is often used for maps of the earth, since it preserves the shape of small surface regions. It also maps circles on the globe to circles on the plane, and in higher dimensions, maps spheres to spheres.

The northern regions are projected onto a plane tangent to the central sphere at the north pole N. A point P on the sphere projects to the point Q where the ray from the south pole S through P meets the tangent plane. This projection is defined for all points P except the south pole S. The projection from N onto the plane tangent at the south pole is defined similarly, as shown in figure 3 in the 3-D case.

Spheres will project as round spheres, but the radius will increase as they approach the projecting pole. When the sphere passes through the pole, it projects as a hyperplane (a straight line in 2-D). However each projection produces only moderate size distortion in the hemisphere about the point of tangency, and gives a good indication of the sphere arrangement there.

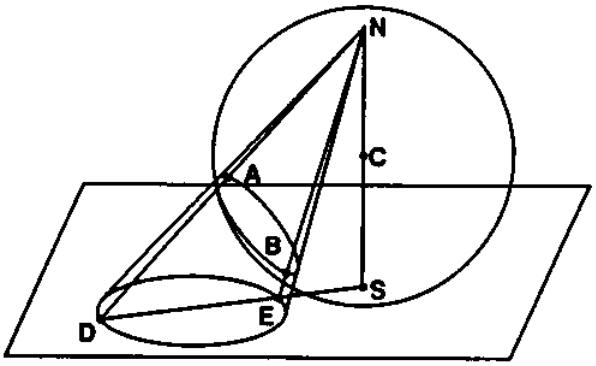

Figure 3. Stereographic projection from the north pole.

The next three paragraphs show why stereographic projection takes circles to circles in the 3-D case, using a geometric argument. The argument generalizes to show that spheres project to spheres in higher dimensions. A proof can also be made algebraically, using analytic geometry. In figure 3, the circle through A and B projects to a rounded curve through D and E, which we want to show is a circle. This curve is the intersection of the tangent plane DS at S with the cone through N and the circle AB. This cone has a circular cross section in the plane of circle AB, but an elliptical cross section in a plane perpendicular to its axis. We will show that the planes AB and DS make equal angles to the axis, and that this makes the cone intersect plane DS in a circle also.

Figure 4 shows a cross section in the plane through N, S and the center K of the circle AB. In figure 3, A is the highest point on the circle AB and B is the lowest, so they both lie in this cross section. The radius from the sphere center C through K intersects the sphere at H, and arcs AH and HB are equal. Since an inscribed angle is half of its intercepted arc, ∠ANH = ∠HNB. Thus the elliptical cone is symmetric with respect to the plane through NH, perpendicular to the section of figure 4. Reflection in this plane takes circle AB to another circle FO, and ∠ABN to an equal angle, ∠NFG. If we can show that the plane of FG is parallel to the projection plane DS, then the curve DE will be a magnified version of the circle FG, and thus also a circle.

To do this it is sufficient to show that ∠NFG =∠NDS. Now since NS is perpendicular to DS, ∠NDS = 90° - ∠DNS. Also, arc AS + arc AN = arc NAS = 180°, and ∠DNS = ∠ANS = 1/2 arc AS, so ∠NBA = 1/2 arc AN = 90° - ∠DNS. Finally, ∠NFG = ∠NBA by reflection symmetry, so ∠NFG = ∠NDS as required.

The situation in figures 3 and 4 indicates that the points D and E lie at ends of a diameter of circle DE, so the center of the circle is at the midpoint of segment DE. Thus the center and radius of the circle can be defined by projecting just the top and bottom points A and B. This is true in higher dimensions also, and this method was used to compute the stereographic projections of the spheres. I produced cross-eyed stereo pairs of the stereographic projection for each hemisphere, as shown in figures 5 to 8, using parallel projection from two slightly different angles. Since the spheres should not intersect, I sorted them from back to front, and used the filled circle subroutine in Xwindows to block out the hidden arcs and lines.

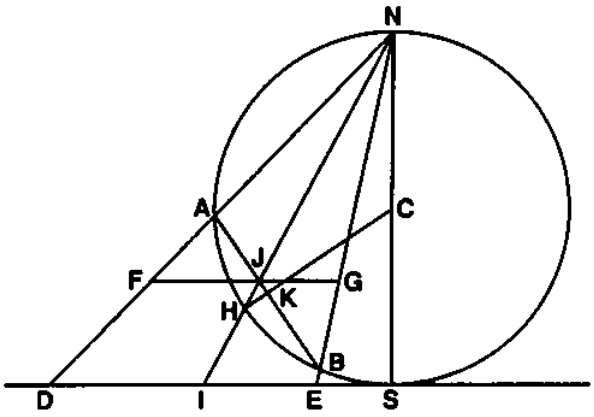

Figure 4. A vertical cross section of figure 3.

An interactive program allowed the user to place a sphere in 4-D, either by giving the center coordinates directly, by giving a relative change from a previous value, or by specifying other spheres that it should touch (particularly useful for generating a close packing). In 4-D, a sphere may be placed to touch four given spheres. There are two such positions, on opposite sides of the hyperplane through the centers of the four given spheres. Therefore, a fifth sphere not on this hyperplane is specified, and the new sphere is placed on the side opposite to this sphere.

Whenever a sphere is positioned, its distances to the other spheres are computed, and any close contacts are reported. The user may also read or write coordinates to a file, rotate the stereographic projections in 3-D, or change the style of the drawings. (Compare figures 5, 6, and 7.) The Voronoi cell construction and volume computation described below may also be requested at any stage.

Figure 5 shows the 24 spheres in the D4 packing. There are six spheres at the vertices of a regular octahedron around the north pole. Next, there are 12 spheres at the equator, arranged as in the FCC packing. These appear in both views, and are larger due to the stereographic projection size distortion. Finally, there are six more spheres in an octahedron around the south pole. The equator is the large dotted circle in each view.

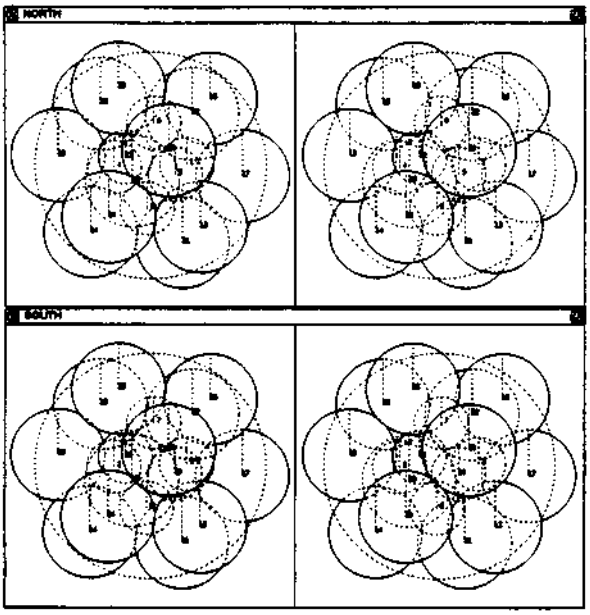

Figure 5. Cross-eyed stereo version of D4 packing.

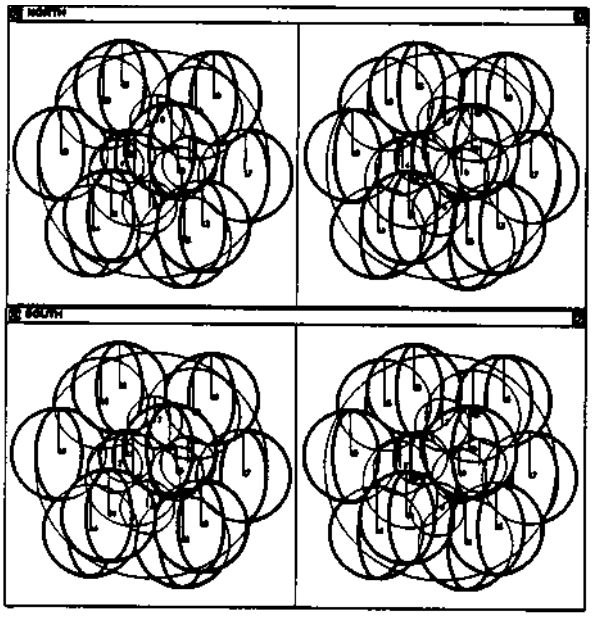

Figure 6. Another style for drawing the Figure 5 spheres.

Figure 6 shows the same geometry drawn in a different style, with a thick profile circle and two arcs at 60 and 120 degrees longitude on the front surface of each sphere, which give the impression of roundness when viewed in stereo.

Figure 7 shows a different configuration, with 10 spheres in the northern hemisphere: number “1”, at the north pole, and nine others, all touching sphere number “1” as well as the central sphere. This packs spheres more tightly near the north pole than in the D4 case. However, when nine additional spheres are placed touching the first nine, their centers lie in the southern hemisphere, and there is only room for one more sphere, number “20”, in the space between them. The spheres were placed by specifying others which they touched, and interactive computer investigation demonstrated that there is no room for a 21st sphere.

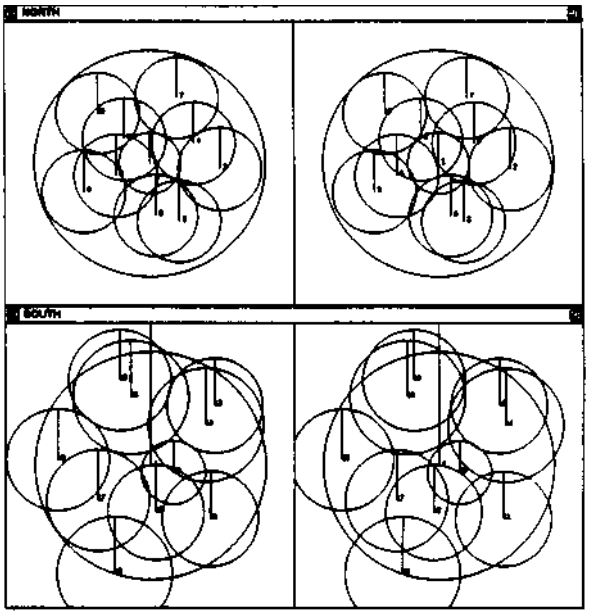

Figure 7. Another packing, with 9 spheres surrounding the north pole.

Rigidity (question b)

In the FCC (or D4) packing, the arrangement of the 12 (or 24) spheres surrounding a central one is tight, so that no sphere can move unless its neighbors also move. (In the regular dodecahedron kissing configuration, the spheres have room to rattle.) Thus any motion must be a concerted one, with all the surrounding spheres sliding at once. This puts constraints on the initial velocity $v[i]$ of each of the 12 (or 24) spheres. Let $p[i]$ be the position of sphere i, with the central sphere at the origin. Then the condition for maintaining tangency to the central sphere requires that

$$v[i] \cdot p[i] = 0.$$

This gives 12 (or 24) conditions for the 36 (or 96) components of velocity. In addition, each surrounding sphere initially touches 6 (or 8) others, and the initial velocities preserve to first order the distances between pairs of spheres in contact, because the arrangement is tight. This gives another equation

$$(p[i] - p[j]) \cdot (v[i] - v[j]) = 0,$$

for each pair (i, j) of surrounding spheres which touch each other. There are $6 \cdot 12/2 = 36$ of these pairs in the FCC case, and $8 \cdot 24/2 = 96$ in the D4 case, giving totals of 48 or 120 equations, respectively. These equations are not all independent, however. The number of independent equations can be found from the rank of the 48 by 36 (or 120 by 96) coefficient matrix. Since the equations are homogeneous and the entries of the matrix are all integers (1,-1, or 0), the rank and a basis for the space of solutions can be found by gaussian elimination with exact integer arithmetic. When I did this calculation in the FCC case, I found that the matrix has rank 32, so the space of solutions is of dimension 36 - 32 = 4. Three of the basis elements can be chosen to be velocities of rigid body rotations, since the rotation group for 3-space has dimension 3. The fourth independent element describes the initial velocities shown in figure 1 for the sliding motion. In four dimensions, the rank of the matrix turns out to be 90, so the solution space has dimension 96 - 90 = 6. Since the group of rotations of 4-space also has dimension 6, each initial velocity solution is the derivative of a rotation, and it follows, as explained in the appendix, that there are no sliding motions. This answers question b) in the negative.

Voronoi cell volume (problem c)

Given a collection of points in n-space, the Voronoi cell C of one of the points P is the set of all positions which are at least as close to P than to any other point in the collection. If Q is another point in the collection, the set of positions equidistant from P and Q is the hyperplane H perpendicular to and bisecting the line segment PQ. This hyperplane H forms part of the boundary of C (unless all positions on it are closer to some other point in the collection). Thus C is the intersection of easily computed half-spaces, and can be constructed by intersecting these half-spaces one by one. Once C is known, it can he chopped up into simplices, whose volume can be calculated using determinants. We first discuss the construction of C, and then the volume calculation. Both calculations were written to work for arbitrary dimension n.

Since C is a polyhedron, it can be defined by lists of faces in each dimension, together with incidence information which gives the boundary of each face as a sum of faces in the next lower dimension. The data structure for a face (or for the whole polyhedron) includes a name, a link in the list of faces of the same dimension, a linked list of boundary face pointers in one dimension lower, and a linked list of vertices.

When a polyhedron is intersected by the half-space on the positive side of the hyperplane H, the equation for H is evaluated and stored with each vertex. Next each edge is clipped by H. If an edge has both a positive and a negative vertex, a new vertex is created where the edge intersects H, and the edge is replaced by its non-negative portion. Then, in order of increasing dimension, faces with only negative vertices are eliminated, and each face F which intersects H is clipped. The negative vertices of F, and the boundary faces of F which have only negative vertices, are eliminated from the appropriate lists. A new "capping" face is added representing $F \cap H$, whose boundary consists of the capping face $G \cap H$ for each boundary face G of F which intersects H.

For a finite collection of points, some Voronoi cells will be semi-infinite polyhedra. In order to handle these, the data structures must be more general. For example, some edges may have zero or one vertex, instead of two. However, such semi-infinite cells will have infinite volume, and will not be used in estimating the packing density. Therefore I wrote the program to handle only finite polyhedra.

This still presents difficulties, because when even a finite polyhedron is constructed as an intersection of half-spaces, the initial few stages will be semi-infinite. l therefore started with a large hypercube, and intersected the hyperplanes with it. If any of the original hypercube faces remain, the size of the hypercube is doubled, and the construction is repeated. After a doubling limit is exceeded, an error is reported. (Finding the face structure of a hypercube in arbitrary dimension n requires some calculation with vertex index permutations.) A Voronoi cell is convex, so it can be recursively divided up into n-simplices, each given by n+1 vertices $v[0]$, $v[1]$, ..., $v[n]$. The volume of such a simplex is $1/n!$ times the absolute value of the determinant of the n by n matrix whose columns are the n vectors $v[1] - v[0]$, $v[2] - v[1]$, ..., and $v[n] - v[n-1]$.

The simplest formula for a determinant involves the sum of $n!$ terms, each a product of n matrix elements. However it can be computed in fewer steps with the LU decomposition. As explained in [3], M is decomposed as the product $L \cdot U$ of two square matrices, where L is lower triangular (all elements above the main diagonal are zero) and U is upper triangular (defined similarly). In addition, the matrix L can be assigned to have all its diagonal elements equal to 1. Then the determinant of M is just the product of the diagonal elements of U.

I used the Crout algorithm (see [3]) to compute L and U, which takes $n^3/3$ multiplications and $n^3/3$ additions. I was able to reduce the arithmetic even further by combining the Crout algorithm with the recursion used to subdivide C. This was possible because the i'th columns of L and U depend only on the columns of M up to column i.

The recursion to subdivide C starts by choosing an initial vertex $v[0]$ of C, and considering a pyramid from $v[0]$ to each face of C which does not contain $v[0]$. For each such face F, an initial vertex $v[1]$ is chosen, and the process is repeated recursively. A chain of the recursion terminates with a list of n+1 vertices for a simplex, with earlier vertices appearing together in more and more such simplices. I rewrote the Crout algorithm to compute M, U,

and L one column at a time, and accumulated the product of the diagonal elements of U during the recursion. Thus the computation for the early columns is amortized over many simplices.

Minimum Voronoi Cell (problem d)

The proof in [1] that the dodecahedron gives the minimum Voronoi cell volume in 3-D uses a volume calculation which works for any solid bounded by planes tangent to the sphere. If we take the collection of spheres touching at the respective points of tangency, we get a generalized collection of kissing spheres, which all touch the central one, but are allowed to interpenetrate each other, so the dodecahedron is also the minimum solution to this more general problem. If the same sort of calculation would work in four dimensions, the Voronoi cell from D4 should also give the global minimum for the generalized problem. However, Warren D. Smith of NEC Research Institute, Princeton NJ, has conducted a search for a maximum volume inscribed polyhedron with 24 vertices on the unit sphere in 4-space. He found the 24 vertices listed in table l, whose convex hull has volume 2.18818. This exceeds the volume 2.0 from the 24 vertices for D4.

0.749954	-0.643197	-0.084056	0.129616
-0.828394	-0.525978	0.180627	-0.066966
-0.190731	0.714818	-0.397429	0.542869
-0.420367	0.043714	0.377460	-0.823957
0.264726	0.358996	-0.289371	-0.846939
0.107640	0.567716	-0.808451	-0.111887
0.681175	-0.247155	0.120147	-0.678587
0.296790	-0.078669	-0.729281	0.611454
-0.536970	-0.077939	0.045792	0.838744
-0.414729	0.118922	0.901507	0.033811
-0.144972	0.934721	0.131967	-0.296420
-0.595497	-0.191546	-0.753519	0.202243
-0.620266	0.250172	-0.524717	-0.526646
0.385563	-0.274729	-0.829731	-0.295655
0.683734	-0.065639	0.705573	0.174258
0.838599	0.512987	-0.168448	0.072261
-0.155567	-0.682066	-0.315419	-0.641167
0.377718	0.425282	0.674618	-0.470483
-0.877111	0.451110	0.124560	0.107982
0.079658	0.629705	0.538647	0.554062
-0.021552	-0.526587	0.616342	0.585119
0.464299	-0.021498	0.032543	0.884819
-0.093366	-0.877397	-0.276253	0.380975
0.062900	-0.701895	0.621751	-0.341779

Table 1.

When l tested the tangent planes at these points with the algorithm of the previous section, l found a volume of 7.97690, which is smaller than 8.0, the volume for the Voronoi cell of D4. More extreme examples may exist. Thus D4 does not give the minimum solution to the generalized problem, though it may still be the best for the strict kissing sphere problem.

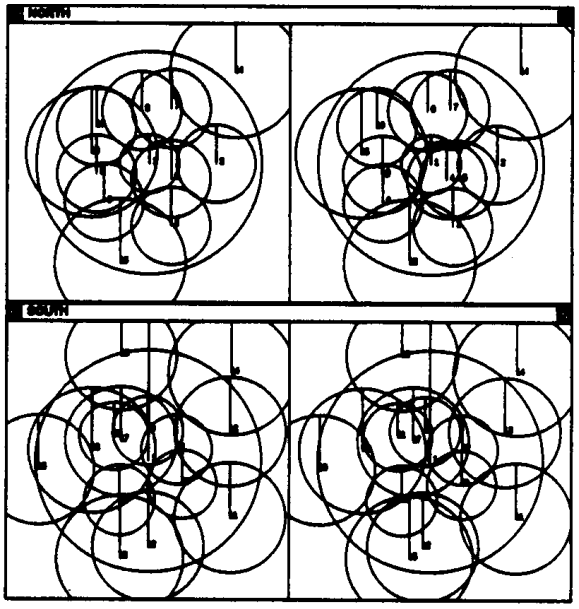

Figure 8. Twenty-four spheres at an equal distance l.09 from a central sphere.

The collection of 20 spheres shown in figure 7 defines a Voronoi cell around the central sphere of volume 9.0974 because of the inefficient packing around the south pole. Perhaps this packing could be improved by moving all the spheres farther away from the central one. I found that if I increased the radius of the central sphere to 1.099965, and kept the radii of the other kissing spheres at 1.0, I could just squeeze in 24 spheres, as shown in figure 8, while keeping the basic arrangement near the north pole close to that in figure 7. (Precisely, the 9 great circles through the north pole and one of the kissing points of the 9 sphere surrounding the north are the same.) The Voronoi cell for the central sphere in this configuration was 9.72727, reflecting the fact that the surrounding spheres are farther away. When they are all moved inwards so that they overlap each other but kiss the central sphere, the volume decreases to 8.15590, which is larger than the D4 solution to the “strict” kissing sphere problem.

Acknowledgments

This research was performed under the auspices of the U. S. Department of Energy by Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48, and was also supported by NSF grant No. DMS-86l562-01 to the University of Minnesota Geometry Supercomputer Project. Questions a), b), and c) were suggested by Wu-yi Hsiang. Gary Rodrigue suggested the LU method for computing determinants. 1 had helpful discussions with Warren Smith who supplied the data for table l. One of the visualization ’91 reviewers suggested figures 2 and 3.

References

[1] Hsiang, Wu-yi, “On the density of sphere packings in E3: Kepler’s conjecture and Hilbert’s 18th problem” preprint PAM-516, Center for Pure and Applied Mathematics, Univ. of Calif., Berkeley (1990)

[2] Conway, J. H. and Sloane, N. J. A., “Sphere packings, lattices and groups,” Springer Verlag, Berlin (1988)

[3] Press, William H., Flannery, Brian P., Teukolosky, Saul A., and Vetterling, William T., “Numerical Recipes in C: The Art of Scientific Computing,” Cambridge University Press, New York (1988)

Appendix

In the analysis of problem b), we showed that if $F(t)$ is an allowed motion of the 24 kissing spheres, then the initial velocity $F'(0)$ is the derivative of a rotation. That is, if P is the 24 by 4 matrix of coordinates for the 24 sphere centers, then there is a family of rotations $R(t)$ so that $F'(0)=R'(0)P$. This means that $R(t)P$ agrees with $F(t)$ to first order, at $t = 0$. But what about the higher derivatives? Since the constraint that the 24 kissing spheres do not overlap is analytic, we will consider only analytic motions $F(t)$, which are defined by a power series near $t = 0$. We will show that any such F is actually a family of rotations of P by repeated use of the first derivative property above. We already know that $R(t)P$ agrees with $F(t)$ to first order, so the motion

$$M(t) = R^{-1}(t)\, F(t)$$

has zero derivative at t = 0. Let

$$G(t) = M(t^{1/2})$$

so that

$$G(u^2) = M(u)$$
$$dG(u^2)/du = 2uG'(u^2) = M'(u)$$
$$G'(u^2) = M'(u)/(2u)$$

and

$$G'(0) = 1/2 \lim ((M'(u) - M'(0))/(u - 0)) = 1/2\, M''(0).$$

Then, since $G(t)$ is a nonoverlapping motion, there is a family of rotations $S(t)$ so that $S'(0)P = G'(0)$. One can check that the family of rotations $R(t)S(t^2)P$ agrees with $F(t)$ to second order.

To continue one step more, the motion given by the equation $N(t) = S^{-1}(t^2)R^{-1}(t)F(t)$ has zero first and second derivatives at $t = 0$. Using the derivatives of $N(t^{1/3})$, and the fact that $F(t)$ is analytic to make the limits work, one can find as before a family of rotations $T(t)$ so that $T'(0)P = 1/6\, N'''(0)$. One can then show that $R(t)S(t^2)T(t^3)P$ agrees with $F(t)$ to third order. If we continue this process to the limit, we can match all the derivatives of F with a family of rotations. Since F is analytic, this matches F itself, so F is a family of rotations of P, and not a sliding motion.